Brief Introduction to Latent Class, Latent Transition, and Growth Mixture Models

Jason T. Newsom

Department of Psychology
Portland State University

Latent Class Analysis

- Latent class analysis (LCA) is a general framework for grouping individuals into probable classes
- Difference from factor analysis: in LCA we are grouping individuals using a set of variables not grouping variables using a set of people
- The number of classes and their interpretation is unknown and must be determined
- There are a variety of variations on latent class models, specific estimation approaches, and applications

General Modeling Approaches

- There are a couple of general approaches that have many things in common, but I will focus primarily one that is implemented within a structural equation modeling framework (available in Mplus; Muthén & Muthén, 1998–2018)

- This approach follows a confirmatory strategy because it can be used with larger predictive models with latent classes as predictors or outcomes ("auxiliary variables", "distal outcomes")

- Other software packages include: poLCA and lcca packages in R, PROC LCA, which is a free macro for SAS (Lanza, Collins, Lemmon, & Schafer, 2007), Latent Gold (Vermunt & Magidson, 2005), and Mx (Boker et al., 2012)

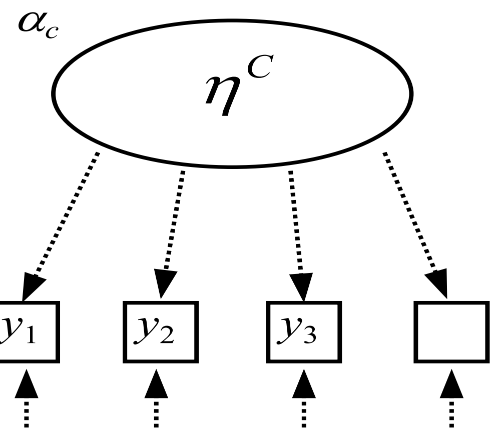
Concepts

- One can think about latent class models within a kind of generalized structural equation modeling framework in which latent classes are latent categorical variables (Muthén, 2001)[1]
- This approach allows latent classes to be used with larger predictive models with latent classes as outcomes or predictors

[1]Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class-latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), New methods for the analysis of change (pp. 291–322). Washington, DC: American Psychological Association.

Concepts

- A set of measures can be used to define a latent categorical variable
- These indicators can be binary (latent class analysis) or continuous (latent profile analysis)



Concepts

- Categories are unobserved or unknown rather than known
- For example, if we have a set of political opinion items (agreement or yes/no) for issues such as abortion, taxes, immigration, crime, global warming, etc.
- From responses, people can be grouped in to some set of homogeneous classes, which might be two (e.g., conservative vs. liberal, three (progressive, libertarian, conservative), or four (libertarian, MAGA, moderate conservative, socialist).
- Contrast with registered or self-identified groups like Democrat, Republican, Green Party, Independent

Concepts

- The number of classes and their meaning must be decided upon based on fit with the data

- Typically several models are estimated, specifying one, two, three, four or more classes, and fit is compared among these models

- Although ideally these could be thought of as nested models, the usual likelihood ratio test does not work well

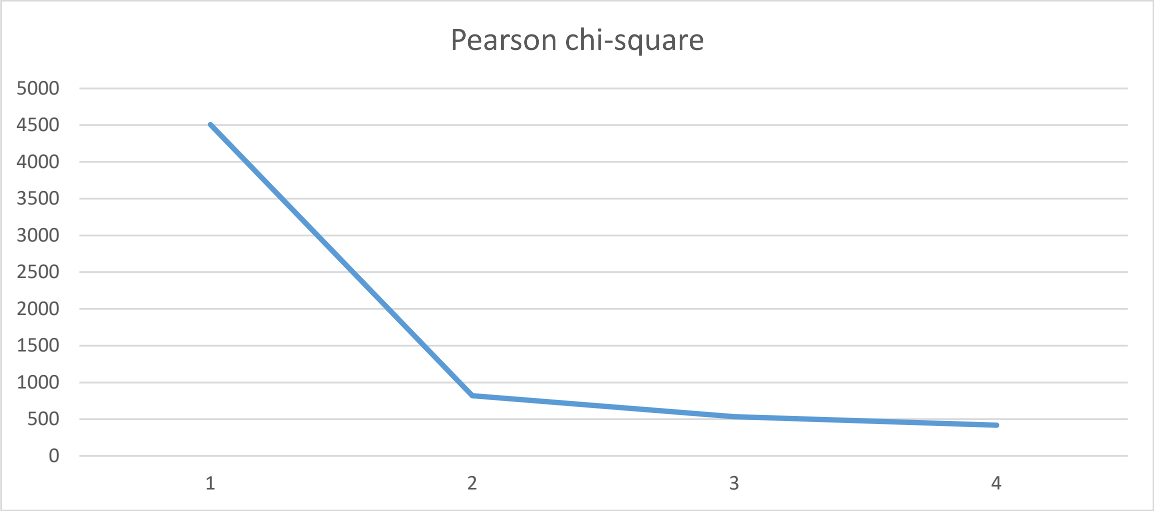
- Sample size adjusted Bayesian Information Criteria (aBIC, ABIC, or SABIC; Sclove, 1987) or bootstrapped likelihood ratio test (BLRT; McLachlan & Peel, 2000;) work best (Nylund et al., 2007)

Example[2]

- Determining multimorbidity classes based on set of eight chronic conditions (e.g., diabetes, heart disease, cancer)

classes	VLMR LR	BLRT p
1	na	na
2	19.962	<.001
3	75.699	0.0507
4	54.781	0.0118

aBIC



Number Classes

[2]Quiñones, A. R., Newsom, J. T., Elman, M. R., Markwardt, S., Nagel, C. L., Dorr, D. A., ... & Botoseneanu, A. (2021). Racial and ethnic differences in multimorbidity changes over time. *Medical care, 59*(5), 402-409.

Posterior Class Probabilities, Response Probabilities, and Entropy

- After choosing the optimal number of classes, several results are of interest

- Posterior class probabilities give the estimated number of cases and proportion that mostly likely belong to a particular class – probable class membership not assigned to class

- Response probabilities (binary) or conditional intercepts (continuous) are obtained for each indicator in each class, which help identify meaning of classes

- Entropy is an estimate of the homogeneity within classes and class separation (higher numbers better)

Example

Posterior Class Probabilities

- For the multimorbidity example, these are the class memberships for 3 classes

```
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES

    Latent
   Classes

       1        2214.06843          0.30509
       2        2719.39759          0.37473
       3        2323.53398          0.32018
```

Example

Response Probabilities

- Response probabilities are conditional probabilities of the value of 1 on the indicator (or conditional means/intercepts in the case of continuous indicators) given a membership in a particular class

Example

Class 1: Multimorbidity
Class2: Hypertension/arthritis
Class3: Healthy

RESULTS IN PROBABILITY SCALE

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
BP14				
Category 1	0.071	0.017	4.128	0.000
Category 2	0.929	0.017	54.136	0.000
DIAB14				
Category 1	0.572	0.040	14.236	0.000
Category 2	0.428	0.040	10.634	0.000
CANCR14				
Category 1	0.714	0.015	47.850	0.000
Category 2	0.286	0.015	19.199	0.000
LUNG14				
Category 1	0.689	0.028	24.544	0.000
Category 2	0.311	0.028	11.062	0.000
HRT14				
Category 1	0.297	0.069	4.276	0.000
Category 2	0.703	0.069	10.119	0.000
STRK14				
Category 1	0.737	0.042	17.732	0.000
Category 2	0.263	0.042	6.316	0.000
ARTH14				
Category 1	0.049	0.014	3.609	0.000
Category 2	0.951	0.014	70.020	0.000
DEP14				
Category 1	0.708	0.038	18.437	0.000
Category 2	0.292	0.038	7.595	0.000
Latent Class 2				
BP14				
Category 1	0.074	0.067	1.116	0.265
Category 2	0.926	0.067	13.905	0.000
DIAB14				
Category 1	0.627	0.267	2.352	0.019
Category 2	0.373	0.267	1.397	0.162
CANCR14				
Category 1	0.797	0.027	30.001	0.000
Category 2	0.203	0.027	7.644	0.000
LUNG14				
Category 1	0.949	0.107	8.884	0.000
Category 2	0.051	0.107	0.481	0.630
HRT14				
Category 1	0.700	0.086	8.105	0.000
Category 2	0.300	0.086	3.476	0.001
STRK14				
Category 1	0.910	0.056	16.248	0.000
Category 2	0.090	0.056	1.616	0.106
ARTH14				
Category 1	0.253	0.034	7.537	0.000
Category 2	0.747	0.034	22.281	0.000
DEP14				
Category 1	0.945	0.017	55.588	0.000
Category 2	0.055	0.017	3.206	0.001
Latent Class 3				
BP14				
Category 1	0.598	0.530	1.128	0.259
Category 2	0.402	0.530	0.757	0.449
DIAB14				
Category 1	0.911	0.037	24.445	0.000
Category 2	0.089	0.037	2.391	0.017
CANCR14				
Category 1	0.774	0.014	53.928	0.000
Category 2	0.226	0.014	15.744	0.000
LUNG14				
Category 1	0.900	0.024	37.198	0.000
Category 2	0.100	0.024	4.152	0.000
HRT14				
Category 1	0.809	0.020	41.243	0.000
Category 2	0.191	0.020	9.758	0.000
STRK14				
Category 1	0.968	0.009	111.104	0.000
Category 2	0.032	0.009	3.695	0.000
ARTH14				
Category 1	0.316	0.059	5.388	0.000
Category 2	0.684	0.059	11.644	0.000
DEP14				
Category 1	0.934	0.017	54.921	0.000
Category 2	0.066	0.017	3.880	0.000

Example

Entropy

- Higher values indicate greater within-class homogeneity and thus greater differences between class
- A quality or reliability of classification measure of sorts

$$E = \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} \left(-\hat{\pi}_{ic} \ln \hat{\pi}_{ic} \right)}{N \ln C}$$

```
CLASSIFICATION QUALITY

     Entropy                                0.363
```

Latent Transition Models

- Latent transition models are a longitudinal application of latent class models – how stable or changing is class membership over time

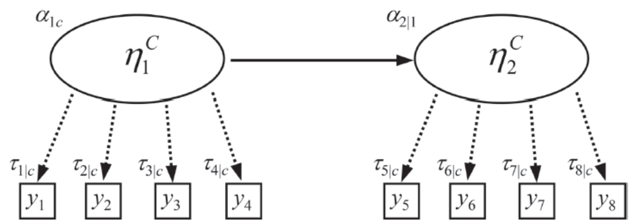


Figure 10.4 Simple Latent Transition Model with Two Time Points.

Latent Transition Models

- Transition probabilities indicate the odds of probable membership in on one outcome class at the later time point given probable membership in a certain class at the initial time point

- With two classes a logistic regression, with three or more, a multinomial logistic regression

$$P\left(\eta_2^C = 1 \mid \eta_1^C = 0\right) = \hat{\pi}_{2|1} = \frac{e^{\alpha_{2|1}}}{1 + e^{\alpha_{2|1}}}$$

Latent Transition Models

- Results from three-class model 1998 predicting three-class model 2014
- Odds ratios: last class of outcome is referent
- Class 1: Multimorbidity; Class2: Hypertension/arthritis; Class3: Healthy

```
Logits for the Classification Probabilities for the Most Likely Latent
Class Membership (Column)

by Latent Class (Row)

             1        2         3

    1     13.788   10.204    0.000
    2     13.407   12.724    0.000
    3     13.204   13.034    0.000
```

Growth Mixture Models

- Growth mixture models combine latent class analysis with growth curve models
- Two general types: multilevel regression approach to growth curve models and semiparametric (latent class growth curve models; Nagin, 1999; e.g., SAS or Stata) or structural equation modeling approach in software that can handle latent class models (e.g., Mplus)
- These are similar in their goals but the Nagin approach does not allow for random slopes within classes
- Valuable for classifying trajectories (e.g., increasing, decreasing, nonincreasing low, nonincreasing high)

Growth Mixture Models

- Number of classes must be select using the same process as that described for latent class models
- Any of the parameters can differ across classes (intercept means, slope means, intercept variances, slope variances, intercept-slope covariance, occasion variances), typically at least intercepts and slope means)
- Often computationally intensive and can be difficult to estimate without convergence problems. Constraining parameters equal across groups can help with estimation difficulty.

Growth Mixture Models

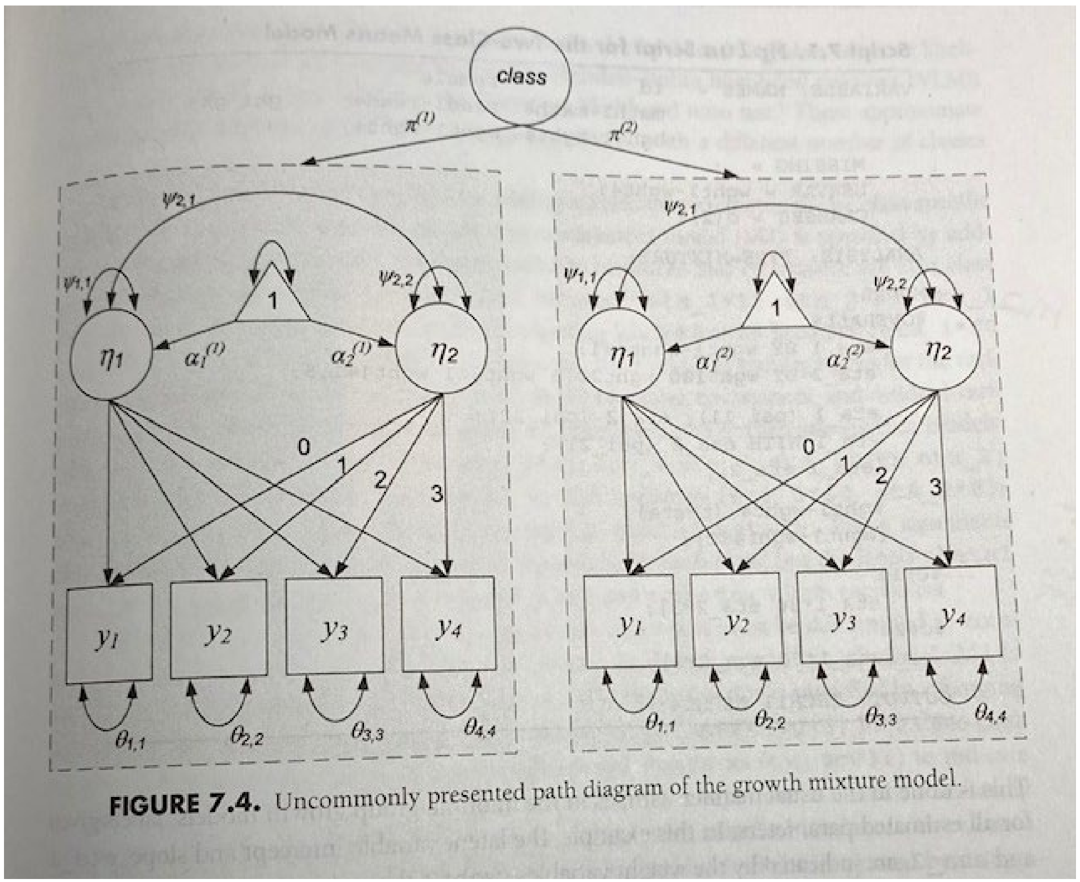


FIGURE 7.4. Uncommonly presented path diagram of the growth mixture model.

From Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications., p. 145

Growth Mixture Models

- Equal variances often assumed but can cause incorrect conclusions about the number of classes or biased estimates (Diallo et al., 2016; McNeish & Harring, 2021)
- May need at least 1,000 cases to make estimation with heterogeneous variances across classes

Example[3]

- We examined changes in multimorbidity over 18 years in mid to late life (biennial)

- Results concluded three classes (low, increasing, high)

Table 1
Class enumeration of latent multimorbidity trajectories (N = 13,699).

Number of Trajectories	BIC	BLRT *p*	Entropy
1	261,884.91	–	–
2	260,956.55	.000	.717
3	260,175.42	.000	.707
4	259,946.25	.000	.703
5	259,741.28	.000	.700

Abbreviations: BIC = Bayesian Information Criteria, BLRT *p* = bootstrapped likelihood ratio test p-value.

[3] O'Neill, A. S., Newsom, J. T., Trubits, E. F., Elman, M. R., Botoseneanu, A., Allore, H. G., ... & Quiñones, A. R. (2023). Racial, ethnic, and socioeconomic disparities in trajectories of morbidity accumulation among older Americans. *SSM-Population Health*, *22*, 101375.

Example

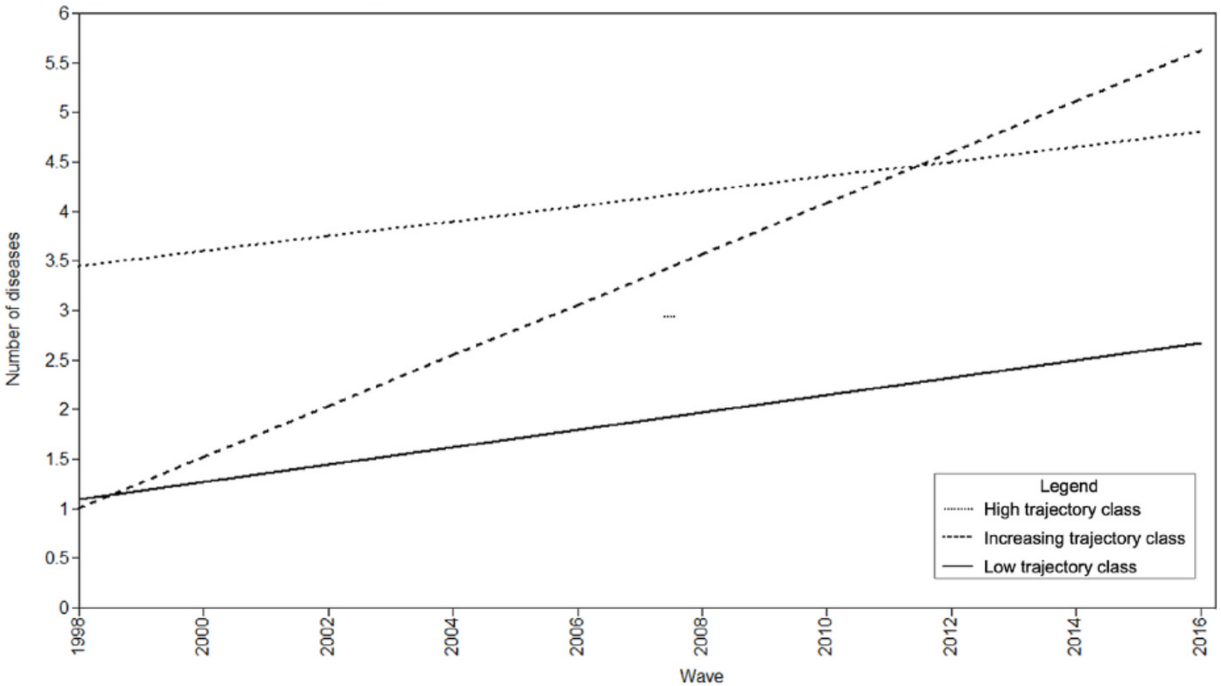


Fig. 2. Graph of the **low**, **increasing**, and **high** trajectories from the unconditional model (N = 13,699).

Example

- Race and ethnicity differences of growth curve class membership

Table 3
Unconditional likelihood of class membership into trajectory for racial and ethnic comparisons (N = 13,439).

Trajectory comparison	Racial/ethnic group	OR	CI	p-value
High versus low				
	NH Black	**2.925**	**2.419, 3.537**	**.000**
	Hispanic	1.238	0.977, 1.567	.077
Increasing versus low				
	NH Black	1.203	0.914, 1.582	.187
	Hispanic	1.252	0.941, 1.666	.123
High versus increasing				
	NH Black	**2.432**	**1.837, 3.219**	**.000**
	Hispanic	0.988	0.721, 1.355	.943

Note. Low – Low trajectory, Increasing – Increasing trajectory, High – High trajectory, NH – Non-Hispanic. NH White is the referent group. This model was not adjusted for additional covariates beyond race/ethnicity. All estimates are unstandardized. Significant effects are bolded.

Example

References

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., …, Brandmaier, A. (2012) OpenMx version 1.3. Retrieved from http://openmx.psyc.virginia.edu.

Diallo, T. M., Morin, A. J., & Lu, H. (2016). Impact of misspecifications of the latent variance–covariance and residual matrices on the class enumeration accuracy of growth mixture models. Structural Equation Modeling: A Multidisciplinary Journal, 23, 507–531.

McLachlan, G., & Peel, D. (2000). Finite mixture models. New York: Wiley.

McNeish, D., & Harring, J. R. (2021). Improving convergence in growth mixture models without covariance structure constraints. Statistical Methods in Medical Research, 30, 994–1012.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class-latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), New methods for the analysis of change (pp. 291–322). Washington, DC: American Psychological Association.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. Psychological Methods, 4, 139–157.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent classanalysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling: AMultidisciplinary Journal, 14, 535–569.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. Psychometrika, 52, 333–343.

Vermunt, J. K., & Magidson, J. (2021). Upgrade manual for latent GOLD basic, advanced, syntax, and choiceversion 6.0. Arlington, MA: Statistical Innovations Inc.

Example

Recommended Readings

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. Psychological Methods, 9, 3–29.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg & M. E. Sobel (Eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 311–359). New York: Plenum.

Collins, L. M., & Lanza, S. T. (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences (Vol. 718). New York: Wiley.

Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications., p. 145

Newsom, J.T. (2024). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge. Chapters 10 & 11

Wang, M., & Bodner, T. E. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. Organizational Research Methods, 10, 635–656.

Thank you!

Please contact Jason Newsom, newsomj@pdx.edu, with comments or questions.